# Agile Big Data Analytics of High-Volume Geodetic Data Products for Improving Science and Hazard Response

Completed Technology Project (2015 - 2017)

## Project Introduction

Geodetic imaging is revolutionizing geophysics, but the scope of discovery has been limited by labor-intensive technological implementation of the analyses. The Advanced Rapid Imaging and Analysis (ARIA) project has proven capability to automate SAR image analysis, having processed thousands of COSMO-SkyMed (CSK) scenes collected over California in the last year as part of a JPL/Caltech collaboration with the Italian Space Agency (ASI). The successful analysis of large volumes of SAR data has brought to the forefront the need for analytical tools for SAR quality assessment (QA) on large volumes of images, a critical step before higher level time series and velocity products can be reliably generated. While single interferograms are useful for imaging episodic events such earthquakes, in order to fully exploit the tsunami of SAR imagery that will be generated by current and future missions, we need to develop more agile and flexible methods for evaluating interferograms and coherence maps. Our AIST-2011 Advanced Rapid Imaging & Analysis for Monitoring Hazards (ARIA-MH) data system has been providing data products to researchers working on a variety of earth science problems including glacial dynamics, tectonics, volcano dynamics, landslides and disaster response. A data system with agile analytics capability could reduce the amount of time researchers currently spend on analysis, quality assessment, and re-analysis of interferograms and time series analysis from months to hours. A key stage in analytics for SAR is the quality assessment stage, which is a necessary step before researchers can reliably use results for their interpretations and models, and we propose to develop machine learning tools to enable more automated quality assessment of complex imagery like interferograms, which will in turn enable greater science return by expanding the amount of data that can be applied to research problems. Objectives: We will develop an advanced hybrid-cloud computing science data system for easily performing massive-scale analytics of geodetic data products for improving the quality of InSAR and GPS data products that are used for disasters monitoring and response. We will focus our analysis on Big Data-scale analytics that are needed to quickly and efficiently assess the quality of the increasing collections of geodetic data products being generated existing and future missions. Technology Innovations: Science is an iterative process that requires repeated exploration of the data through various what-if scenarios. By enabling faster turn-around of analytics and analysis processing of the increasing amount of geodetic data, we will enable new science that cannot currently be done. We will adapt machine learning approach to QA assessment for improving the quality of geodetic data products. Furthermore, these types of analytics such as assessing coherence measures of the InSAR data will be used to improve the quality of the data products that are already being used for disasters response. We will develop new approaches enabling users to quickly develop, deploy, run, and analyze their own custom analysis code across entire InSAR and GPS collections. Expected Significance: To improve the impact of our generated data products for both the science and monitoring user communities, quality assessment (QA) techniques and metrics are


Agile Big Data Analytics of High-Volume Geodetic Data Products for Improving Science and Hazard Response

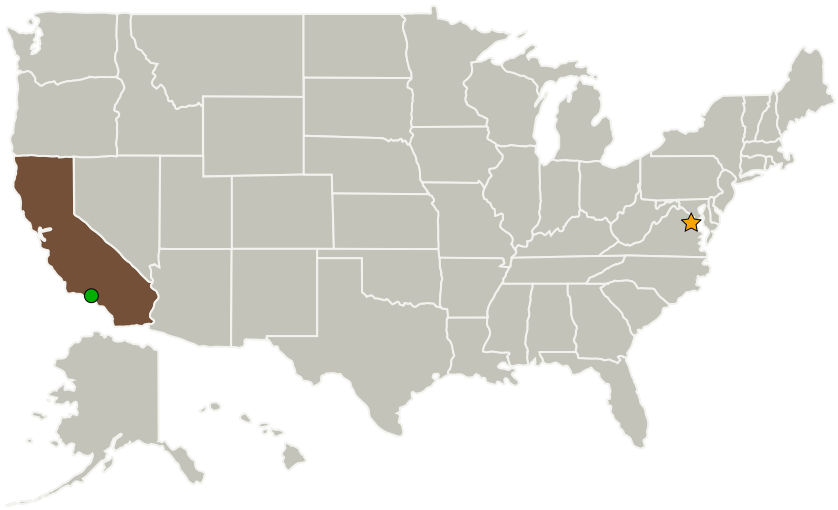
## Table of Contents

# Agile Big Data Analytics of High-Volume Geodetic Data Products for Improving Science and Hazard Response

Completed Technology Project (2015 - 2017)

needed to automatically analyze the PB-scale data volumes to identify both problems and changes in the deformation and coherence time series. Automated QA techniques are currently underdeveloped within the InSAR analysis community, but have already become much more strategically important for supporting the expected high data volumes of upcoming missions such as Sentinel, ALOS-2, and NASA-ISRO SAR (NISAR) and high-quality science and applications. The science data system technology will also enable NASA to support the high data volume needs of NISAR in addition to the analysis of the data products.

### Primary U.S. Work Locations and Key Partners



| Organizations Performing Work | Role | Type | Location |
|---|---|---|---|
| ★NASA Headquarters(HQ) | Lead Organization | NASA Center | Washington, District of Columbia |
| ●Jet Propulsion Laboratory(JPL) | Supporting Organization | NASA Center | Pasadena, California |

| Primary U.S. Work Locations |
|---|
| California |

## Organizational Responsibility

**Responsible Mission Directorate:**

Science Mission Directorate (SMD)

**Lead Center / Facility:**

NASA Headquarters (HQ)

**Responsible Program:**

Advanced Information Systems Technology

## Project Management

**Program Director:**

Pamela S Millar

**Program Manager:**

Jacqueline J Le Moigne

**Principal Investigator:**

Hook Hua

**Co-Investigators:**

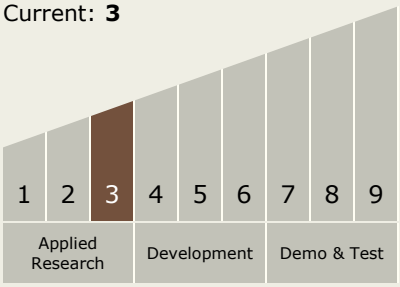
Zhen Liu
Gian Franco Sacco
Piyush S Agram
Carlos J Velazquez
Sang-ho Yun
Paul R Lundgren
Brian D Bue
Susan E Owen-mccollum
Angelyn W Moore
Gerald John M Manipon
Mark Simons
Eric J Fielding
Tom G Farr
Paul A Rosen

# Agile Big Data Analytics of High-Volume Geodetic Data Products for Improving Science and Hazard Response

Completed Technology Project (2015 - 2017)

## Technology Maturity (TRL)

Start: **3**
Current: **3**




## Technology Areas

**Primary:**

- TX11 Software, Modeling, Simulation, and Information Processing
  - TX11.4 Information Processing
    - TX11.4.1 Science, Engineering, and Mission Data Lifecycle

## Target Destination

Earth